The Dark History of HathiTrust

Alissa Centivany
Faculty of Information & Media Studies
Western University
acentiva@uwo.ca

Abstract

This research explores the ways values, power, and politics shape and are shaped by digital infrastructure development through an in-depth study of HathiTrust's "dark history," the period of years leading up to its public launch. This research identifies and traces the emerging and iterative ways that values were surfaced and negotiated, decision-making approaches were strategically modified, and relationships were strengthened, reconfigured, and sometimes abandoning through the process of generating a viable, robust and sustainable collaborative digital infrastructure. Through this history, we gain deeper understandings and appreciations of the various and sometimes surprising ways that values, power, and politics are implicated in digital infrastructure development. Shedding light on this history enables us to better contextualize and understand the affordances, limitations, and challenges of the HathiTrust we know today, better envision its range of possible futures, and develop richer appreciations for digital infrastructure development more broadly.

1. Introduction

Digital infrastructure ("DI") undergirds the platforms, applications, tools, and systems that are increasingly ubiquitous, indispensable, and inseparable parts of life. In contrast to the more public-facing interfaces they support, DI operates beneath the surface, collecting, organizing, and processing data in ways that are difficult to observe and, in many cases, understand and critically evaluate.

This work contributes to understandings of the roles of values, power and politics in DI development through a qualitative study of HathiTrust ("HT"). In 2008, HT was introduced to the public as a shared digital repository ("SDR") jointly launched by the twelve-university consortium known as the Committee on Institutional Cooperation ("CIC") and the eleven university libraries of the University of California System [9]. Emphasizing shared values around information preservation and access and shared traditions around institutional cooperation, HT sought to combine, coordinate and leverage the distributed, independent digitization efforts of its members in the creation of a new DI supporting the "collective collection." In the eight years that have passed since its launch, HT has evolved far beyond these origins. Today, HT has over one hundred institutional partners working cooperatively to sustain and innovate on a DI supporting a growing corpus that, as of this writing, contains over fourteen million digitized print volumes.

These snapshots of HT do not, however, reflect or reveal much about *how* or *why* it came to be or came to become *this* HT. This research describes some of these processes through a telling of HT's "dark history"— the years HT's progenitors spent behind closed doors gestating the digital infrastructure. As with DI development more generally, HT emerged through iterative negotiations, demonstrations and challenges of power, and political posturing and participation. Through this history, we gain deeper understandings and appreciations of the various and sometimes surprising ways that values, power, and politics shape and are shaped by technical, social, and legal/policy concerns in DI development. Shedding light on this history enables us to better contextualize and understand the affordances, limitations, and challenges of the HT we know today and better envision its range of possible futures.

This paper begins by reviewing relevant prior work, drawing primarily on digital infrastructure, digitization and digital library literatures and describing the research methods used. Focus then shifts to HT's emergence, organized around three key moments or turning points in its development where

URI: http://hdl.handle.net/10125/41440
ISBN: 978-0-9981331-0-2

the interplay of values, power and politics proved determinative in the outcome: (1) the decision to join Google's mass digitization project ("MDP"), (2) developing the initial digital infrastructure that would become the technical backbone of HT, and (3) fleshing out the social and political dimensions of HT as a semi-autonomous collective organization operating beneath a persistent partial institutional umbrella. The paper concludes by reflecting on the spectrum of ways values, power and politics influenced the emergence and evolution of HT and briefly noting possible implications for HTs future and the future of digital infrastructures more broadly.

2. Related Literature

This research draws upon digital infrastructure, digitization, library and information science, and organizational sensemaking literatures. In combination, this prior work offers helpful insights into current understandings of DI in the library digitization and signals potential gaps in understandings with regard to the roles of values, power and politics.

Sociotechnical systems and infrastructure literatures provide an overarching guiding perspective for this research. The work of Hughes [13] and Bijker [1] are instructive in their emphasis on the social construction of technology and the in-depth descriptive methods used to tease out and foreground the multidimensional, dynamic, and mutually constitutive web of role of values, power and politics in infrastructure development. Echoing observations made by Kling [16], Edwards and colleagues hone in on some of the particular challenges and tensions slow-moving, self-preserving institutions like libraries face when they attempt to translate their deeply engrained traditions, practices, and values to a new digital environment: "Transformative infrastructures cannot merely be technical; they must engage fundamental changes in our social institutions, practices, norms and beliefs as well" [6:13]. The work of Star [19], Ribes [18], Bowker [2] and others offer insights useful for conceptualizing scale — in terms of size, time and zone of influence — in studies of infrastructure development, reminding us that DI like HT do not spring up as *de novo* fully fleshed forms but rather draw upon and interoperate with much older information and communication practices, norms, and technologies and therefore their study demands sensitivity to the "long now" of DI development.

Law, library and information science have also explored important aspects of DIs including, most notably, risks and affordances of large-scale digitization efforts (e.g. Google's MDP) and associated public interest and social justice implications. For example, Vaidhyanathan discussed potential of the MDP in light of a copyright disequilibrium wrought by new digital technologies and hypothesized that a hasty over-reliance on fair use would not only risk derailing the MDP but could significantly undermine future library digitization efforts as well [21, 22]. Grimmelmann has written extensively on the (ultimately unsuccessful) Google Books Settlement and the dangers associated with concentration in the market for digital access to print materials particularly when much of the material is out-of-print [7, 8]. Numerous library and information science studies have sought to position Google and library digitization projects in relation to each other using a variety of values and metrics. Problems and challenges associated with quality, integrity, and access have been addressed in the context of meta-data, preservation, and search [3, 4, 5, 22]. Citing the overwhelming discourse and rhetoric about the relative "open vs. closed" nature of many digitization projects, Leetaru undertook a comparative analysis of the digitization efforts of Google and the Open Content Alliance finding, in practice, that distinctions between open and closed may be more superficial than commonly assumed [17]. Noting the power of knowledge infrastructures to differentially shape, generate and distribute knowledge and justice, Hoffman has conducted a number of studies that describe and critique the MDP on the basis of its negative implications for gender equality and concerns around self-respect, finding that these interests had been promoted by traditional library practices but did not appear to receive adequate support or protection under Google. [10, 11, 12].

Although he did not address digitization or libraries specifically, Weick's work on organizational sensemaking processes provides both theoretical and methodological guidance for identifying and making sense of the ways that values, power and politics factor into the social construction of digitization and DI [23, 24]. Weick stresses, for example, that sensemaking is the primary site where meanings materialize that inform and constrain organizational identity and action [23]. In particular, important linkages are drawn between action (what Weick calls "behavioral commitments") and processes of post-hoc rationalization and justification. Decision-making and sensemaking are entangled in dynamic and continuously evolving processes of social interaction that, over time, become more ordered, stable, and resilient to criticism. Jones elucidated many of these processes in the context of libraries

and library digitization noting, in particular, the importance of naiveté in jump-starting difficult projects and the eventual, almost centripetal return to domain expertise as a means of bringing the projects to fruition and (back) into alignment with traditional library goals, values, practices, and expectations [14, 15].

Informed by these rich and synergistic literatures, this study describes the various way that values, power, and politics shaped and were shaped by the emergence of HT filling some of the existing gaps in understanding by providing detailed descriptive linkages to organizational sensemaking and decision-making processes.

3. Methods

This research seeks to contribute to understandings of the ways that values, power, and politics shape and are shaped by emerging DI through a qualitative study of HT's emergence and evolution. The primary data for this study were generated from in-depth semi-structured interviews with individuals involved in HathiTrust's development. In total, thirty-two participants were interviewed for this study representing sixteen difference institutional/organizational affiliations. The majority of participants were directly involved in HT but several individuals with competing and/or marginalized interests were also interviewed as were outside individuals with expertise on digitization and copyright law but no formal association with HT. Across the different institutions represented, participants' roles varied and included: current and former university provosts, university librarians, chief information officers, librarians and staff, and advisors, employees, and/or members of HT.

Data coding and analysis followed an iterative, inductive approach. As patterns and themes emerged from the data, interview questions were refined to reflect new considerations and points of possible controversy. A process of member checking was used to further test emerging theories, ensure high-quality reporting, and reorganize and refine themes, patterns, and findings as they emerged. Findings of this study are organized as a diachronic narrative using a storytelling approach. Key observations and analytic reflections are interwoven into the description rather than pulled out as a separate discussion section. The concluding section of this paper does, however, briefly summarize and synthesize key findings.

4. The Dark History of HathiTrust

The story of HT's emergence is organized around three key moments or turning points: (1) the University of Michigan's ("UM") decision to join the Google's MDP, (2) developing the initial digital infrastructure that would become the backbone of HathiTrust, and (3) fleshing out the organizational and institutional aspects of HathiTrust prior to its launch. Each turning point is discussed in turn.

4.1. UM-Google Partnership

> *One of the ways the law gets changed is that it gets broken.*
>
> –co-creator of HT

Although its official launch was not until the fall of 2008, HT's origin story began many years earlier when, during a visit to his alma mater in 2002, Google co-founder Larry Page met with librarians at UM to discuss a possible joint digitization venture. From the start, the MDP was deeply contentious. Murmurings of the project sparked wild speculation, vehement commentary, and strident debate amongst a variety of stakeholders. Objections were levied on the basis of copyright law and policy, economic grounds, access and quality of information issues, how the project might affect traditional library values and practices, and myriad social justice concerns. Given this background context, a reasonable jumping off point might be to ask: How does a traditionally risk averse institution like UM decide to undertake such a politically risky, potentially costly, and legally precarious activity as digitizing its entire (roughly six million volume) print library? Several key patterns of justifications emerged through the interviews conducted for this study.

4.1.1. Digitization is inevitable. Mass digitization was not seen as a goal but a given. Participants were not grappling with *if* but rather *when* and *how* digital conversion of the print library would happen. One of the librarians at UM explained, "For libraries and librarians it's as if digitization is written into our DNA. It is what we have to do."

This technological determinism was widespread amongst many in the research library community but it was not universally adopted by the broader community of stakeholders. For example, some were concerned that projects like the MDP might undermine the livelihood of authors and damage the knowledge economy. Well-respected research has lent credence to the tendency and associated risks of conflating technological progress with progress more generally, particularly when a new technology seems to ignore or fail to accommodate key aspects of the

social environment in which it operates [16]. When there is a mismatch or imbalance between technological change, social norms, and shared expectations and practices, technological "progress" can have a paradoxically deleterious effect on existing social relations and structures.

By in large, decision-makers at UM did not find those sorts of arguments compelling. A key administrator at UM who played a central role in forging the UM-Google partnership explained:

The fact that the Google Library Project causes some people to grow concerned about their livelihood is ultimately a moral argument, not an economic one.

Concerns that mass digitization would undermine existing business models that have enabled some members of the literary and publishing world to flourish economically was not, without more, a compelling justification for resisting change.

4.1.2. Digitization is moral. While the purported moral arguments in support of preserving the status quo were dismissed as invalid bases for rejecting the MDP, a moral argument of a different sort was advanced as a justification for the decision to join the MDP. Several HT progenitors reflected on the "strong belief in the inherent rightness" of digitizing books so they might become more accessible to society. In addition, participants emphasized a utilitarian justification saying that, as a matter of principle, we should not permit the interests of the few to hold back the progress of society as a whole simply because they feel entitled to, have grown accustomed to, or have become dependent on the continued enjoyment of the benefits that accrued to them under an old or outdated regime. UM's Chief Librarian explained:

Goddammit, I want there to be a mechanism where almost everybody in the world has access to almost everything that has ever been published in electronic form at zero marginal cost, perhaps with some subscription fee, but a fairly small one. That is what I think the world ought to look like. For academic work, I think that marginal cost and the subscription fee should probably both be zero. The Google project showed me a feasible path to get there, not a complete path, but the starting point. Let's digitize a whole bunch of stuff so that all that prevents it from being available in the way I'd like it to be available is law and custom. I was optimistic that if we, as a society, have valuable assets, then we, as a society, will figure out how to use them. That was the utopian goal.

These sentiments reflect a shared ideology and set of core values held by key decision makers at UM that drove the decision to join the MDP.

4.1.3. Joining the MDP is pragmatic. Large-scale digitization efforts had been undertaken long before the MDP but these efforts were often plagued by a host of recurring challenges. In particular, projects were often swallowed by constant budgetary pressures and the endless creep of technological obsolescence. By offering to cover virtually all of the costs, complete the project on an extremely fast timeline, and provide some technical reassurances in the form of batch updates and other modest maintenance support Google's proposal ameliorated many of these legacy challenges.

Partnering with Google had pragmatic appeal but HTs progenitors were not convinced that the MDP would succeed. In fact, it was not obvious at the outset what "success" even meant. A co-creator of HT recalled:

We didn't have everything all figured out from the get-go. We knew that this was a great opportunity and we wanted to seize it but we weren't exactly sure what we were going to end up doing with the scans.

Google's financial and technological support, and its engineering throughput, was a leap in the right direction. With a long history of stalled and failed digitization projects fading in the rearview, participants appreciated the pragmatic appeal of a partnership with Google. Google might not guarantee success, whatever that might mean, but it might effectively ensure that this digitization project grows too big to fail.

4.1.4. Joining the MDP adds reputational value. UM's decision to partner with Google was also motivated by a sense that doing so would add reputational value to the institution and, by proxy, to the state. A senior administrator involved in negotiating the UM-Google agreement said:

There was a very strong feeling of Michigan exceptionalism on the part of key players that this is the kind of thing that Michigan does and we should do it. The bravery of UM's President was really laudable. I don't know whether she herself really thought it through but she was basically unafraid. The digitization project resonated with her. It was a risk she was willing to take. She said, 'We're going to go ahead and do this. We're going to partner with Google. We're going to scan all these books. We're going to create this thing.' If you were trying to identify a signature of her presidency, I think this is it.

Partnership with Google, one of the world's most dynamic and innovative companies, bolstered UM's sense of exceptionalism and fed into its unique role and position vis-à-vis the economic well-being of the State.

4.1.5. Joining the MDP as a form of advocacy. The decision to join the MDP was also an exercise of advocacy around copyright law and policy. A senior administrator during the Google negotiations who now heads an academic research library said:

I argued in favor of partnering with Google because it was a move that would force theories. Either people would be silent about it and they would be okay with it or it would force a fair use case that would be on favorable terms for us, assuming we did it right. I remember being very concerned that we either use fair use or we lose it. We were looking at the question prospectively rather than just reactively. Short of licensing something, there is no way to guarantee you won't become a test case for fair use. The only way that you can determine that your use was, in fact, definitively a fair use, is to have a judge tell you that. Part of the challenge around copyright cases is, for the most part, publishers pick cases that they think they will win, and then use those decisions to narrow the scope of fair use. And the Google Library Project felt to me, at least intuitively, like … Man, if we're going to have a discussion about fair use then this is the project to have a discussion of fair use around.

A co-creator of HT shared in that sentiment:

This is probably the showdown that we've all known had to happen. And if we lose, it's not over. And if we win, it probably is over. I didn't ever hear it said but I think there were quite a few people who thought that this is the last chance for people who are really opposed to us digitizing the stuff at all to prohibit us from doing that.

As a land grant institution, UM would likely enjoy some immunity against monetary damages for copyright infringement but those protections did not weigh heavily on the decision of whether or not to partner with Google. A senior administrator said:

We wanted to have the fight on the terms of the fight not because we have sovereign immunity and can't be held liable for infringement. Sovereign immunity really served as a safety valve. In the event that everything went down in flames at least they couldn't get damages.

The potential copyright risks dissuaded a number of institutions from joining the MDP and, of those that did join, the majority avoided digitizing works well-within copyright. By contrast UM adopted an aggressive approach, digitizing its entire library; roughly two-thirds of its approximately six million volume collection was believed to be in-copyright. This choice was partially motivated by a desire to advocate for fair use on behalf of libraries and library digitization efforts.

By breaking the UM-Google partnership down into its key justifications we can begin to see some of the various subtle and overt overlapping ways that values (library digitization is part of our DNA/ digitization is moral), power (UM-Google agreement reflected a strategy/pragmatic partnership/UM exceptionalism) and politics (digitization as copyright advocacy/first-mover advantage) played in HT's origin story. Once the decision to partnership with Google had been made, a new host of opportunities, challenges, and tensions emerged.

4.2. Solving an Instrumental Problem

Google scanned the bulk of UM's library in a leased industrial facility on the outskirts of Ann Arbor. Nearly all aspects of the scanning project — the precise location, the process, the technologies used — were kept strictly confidential even from key UM personnel. Google collected truckloads of books, drove them offsite for scanning, and returned them to the library in perfect order, ready for reshelving. The average turnaround time for a given book was approximately one week and, at its height, Google scanned approximately 30,000 volumes from UM's library each week. As a point of reference, it took the most aggressive and technologically advanced library digitizers a decade to scan less than what Google was able to scan each week.

4.2.1. The Initial DI As scans started flooding in, UM realized it needed a place to put them and so it funded and created an initial DI relatively quickly. A senior information officer at UM called the resulting DI a "forcing function of the thing itself." Almost as quickly as it was created, participants became increasingly concerned that the DI did not provide adequate robust security assurances:

Everyone knew that, to do it responsibly, there had to be a second instance located offsite so that problems that hit you aren't likely to hit them.

It was only after UM had its digital back-up copy of the library, and had built a DI to support it, that it realized it was technically and organizationally under-equipped to deal with the instrumental challenges raised by this new DI. Recognizing that as more partners joined the MDP the need for a secure, trusted, digital repository would grow within the broader research library community, UM hoped it could leverage its initial DI to attract the partners it

needed to fund a much-needed second instance at another institution.

A co-creator of HT realized early on that the optimal solution was a single high-quality DI, funded, supported and shared by additional library partners. The lead architect of the UM-Google partnership and co-creator of HT reflected:

The infrastructure had to be done right. It had to be done in a way that people looked at it and said to themselves, 'This is something we can't not do, but we can't afford to do it on our own and we don't need to do it on our own. We can partner with these guys and it will get taken care of.' If every institution tries to do their own version, it won't be done well. But if we have a single infrastructure, we can do it at a high quality and we can afford to bring in other people. Michigan is already supporting this thing quite well, we just need another instance somewhere else.

A DI initially built to solve an instrumental need of a single institution was now being positions as a central node of a far more expansive, collaborative DI — a shared digital repository ("SDR") — that would serve the common needs and interests of the library community.

4.2.2. Values, Power & Politics in Creating the Second Instance UM turned first to its affiliates in the CIC for support in creating the newly reenvisioned SDR. Reflecting on the social and political capital built up within the consortium, a senior information officer said, "We're good at sharing with each other and building things together. We recognize the advantages of economies of scale." The CIC seemed to be on board in principle but the creation of a SDR was not a high priority for its membership. A senior administrator at UM recalled:

There was no urgency within the CIC about this and, as a result, discussions about the creation of a CIC SDR were vague and moving quite slowly. What would the shared digital repository be? Would it be a CIC project? Would it be a project of some university? Were there other universities involved? Would it be a project of a consortium of universities? How are we going to determine the governance, write the bylaws, and so forth?

A rift characterized by many involved in the negotiations as a "clash of cultures" began forming between technologists and librarians at the various CIC institutions. From the librarians' perspective, their hesitant, slow-moving, detail-oriented decision-making process reflected a culture of collectivism and egalitarianism that was integral to the identity of librarians and which libraries had thrived upon for centuries. The approach reflected a sense of the gravity of their professional responsibility and respect for the status of libraries and librarians in society as the trusted stewards of our shared cultural record. From the technologists' perspective, however, the librarians were "pecking this thing to death." A co-creator of HT who straddled the line between librarian and technologist referred to the CIC discussions as a "Zeno's paradox" whereby the task of creating the SDR was being broken down into an infinite number of smaller tasks, effectively rendering completion of the ultimate goal impossible:

We were 99% of the way there but the rest of the way was very clearly going to be something that we weren't going to be able to accomplish because everybody was splitting that last 1%. This was supposed to be the meeting where we made the final commitment! Instead we had library directors saying, 'Yeah, it seems kind of pricey, maybe we shouldn't have two copies of this. The redundancy thing gains us something but we can save money if we don't do that.' But we at Michigan had already committed to that path! It was very clear to us that we needed to have two copies and a back-up to make it viable.

UM needed the SDR to move forward but it did not have the necessary funding to do it on its own. The CIC had funds but was paralyzed by the details. Negotiations were stuck and participants at UM urgently believed they needed to find a way to move things forward.

4.2.3. A Charmed Relationship Saves the SDR Less than twenty-four hours after negotiations stalled with the CIC, UM had its solution. A senior administrator at UM reached out to a friend and CIO at Indiana University ("IU") and, through a couple of brief phone calls over a matter of hours later, the two institutions had negotiated a deal to jointly fund the SDR. The CIO at IU recalled:

I got a call from the CIO of Michigan saying, 'Our Librarian is going to call you because the CIC librarians are really struggling to figure this out.' Then Michigan's Librarian calls while I'm changing planes in Chicago. He knew that I didn't have a lot of time and he said: 'The shared digital repository governance is fucked. This is not going to happen. I can find about $600,000 per year at Michigan. Can Indiana find about $300,000 per year? We'll tell the CIC that we're going to sort this thing out — we'll be the operators of the shared digital repository and the CIC can be its first client. And down the line, we'll get this moved to something else, but this is the only way to get it done.' I said, 'Well, I'm very intrigued. Just let me consult my Librarian. By the next morning my Librarian had gotten the $300,000 per year and I had squared things away with general

counsel. By noon the next day, I called Michigan back and said 'Indiana is in.'

UM and IU would move forward with the SDR on their own without the rest of the CIC. IU agreed without hesitation to defer to UM on all technical, administrative and other decisions related to the project. As IU's CIO recalled:

I told my guys in research technologies, 'Go do whatever Michigan wants.' And they stood up and literally turned that thing on in 30 or 60 days. And I have to credit the strength of Indiana University's IT organization because that was a bit of a countercultural moment in higher education. In higher education, even in administrative and staff positions, everybody gets a vote and everybody gets a say and you have to reach agreement on things."

Participants at UM and IU credited their "charmed relationship" for the quick decision-making around the SDR:

The charmed relationship isn't structural but personal. We have a lot of personal connections of people who have confidence in each other and in creating good outcomes together. We could jump out into the unknown, without everything figured out in advance, and trust that we would both make smart decisions and solve the obvious emergent problems together.

Shared values, practices, political temperament, and attitudes toward the exercise of power contributed to the "charm." In particular, participants cited:

- Common organizational temperament: "*Both institutions have people in key leadership positions who were more interested in making things happen. Not just studying it, but making it happen.*"
- Close personal and professional bonds amongst senior administrators: "*We are kindred spirits and we complement each other.*"
- Shared attitudes toward advocacy: *"We share the sense that great public research universities have to act now or risk becoming less relevant. That is what drives us."*
- History of successful collaborations including the Sakai learning management system that has been adopted by over 350 colleges and universities around the world: "*Institutions feel like they have to be able to answer every possible foreseeable question before they take the first leap. And so that reservoir of personal capital really helps a lot."*

These and other factors enabled UM and IU to reach a near-frictionless agreement in the creation of a SDR, a second instance of the initial DI that would, in time, ultimately become HT.

4.2.4. Dropping the SDR Bomb

When UM and IU returned to the CIC the following day and announced their intention to create the SDR on their own, it sent shockwaves through the room. One CIC participant recalled:

Oh my God, one day, the CIC is going to do this and the next day, it's just Michigan and Indiana. You can imagine, I mean, whoa, that was like, 'Hey, what happened here?!' It was a bomb!

Another CIC member reflected:

Librarians have a very collectivist culture and for someone to break out and do something this way was not only debatable as a strategy, it violated cultural norms of how librarians tend to do things! And it violated the governance structure of the CIC!

UM's Librarian explained his role vis-à-vis the CIC in the following way:

I was something of a bull in a china shop. I hadn't been a University Librarian for very long. I didn't know the secret handshakes. I was a former Provost. I think I was a suspicious character in the CIC and I think that actually served the whole project well. I tried to be friendly, and we did give a lot, but I was unwilling to be hamstrung by the norm of unanimity that meant so much to my CIC colleagues.

The UM librarian who made the actual announcement concerning the SDR at the CIC meeting recalled:

I said, 'Indiana and Michigan are going to cover the entire costs between the two institutions and if the CIC institutions want to come in now, they can be secondary partners and will pay for part but will not have a seat at the table in the same way.' And there was a catastrophic falling out. One of the library directors turned his back on the table. Literally turned his back to me. Lots of people were very unhappy about it.

Notwithstanding the fallout within the CIC resulting from the SDR announcement, UM and IU continued to push ahead with their plan. Key participants from both institutions met in Indianapolis to discuss strategies for moving ahead with the shared DI. Again, in almost frictionless decision-making the group chose a name for the repository — "HathiTrust," identified a strategy for getting buy-in from additional institutional partners, sketched out basic details for what the repository should look like and how it should operate, and agreed on which aspects of the project could be shelved until some future date … all in a day's work.

4.3. Creation and Launch of HathiTrust

Now that the SDR solved the problem of the second instance of the initial DI, focus shifted toward

how UM and IU might navigate the organizational and institutional fallout and begin to build consensus and partnership once again around the DI. A senior administrator at UM and co-creator of HT reflected that, once the instrument problem had been solved,

My first reaction was 'What will all of the people who were involved with this do? Well, they'll hate us. They'll hate Michigan. Anybody we try to bring in will hate us because we're so hegemonic. So I wasn't worried about the technical side. Michigan and Indiana had that covered. I was worried about the organizational side.

One of the ways that UM and IU sought to diffuse some of the backlash was to assure the CIC that, if they decided to join they would be held out to the public as a founding member:

We ultimately gave them a seat on the board and on the executive committee, and that turned into two seats in time. So I think they've gotten everything they would have gotten, but the bomb was the thing that caused them to move forward.

In addition to making amends with the CIC, UM began working on bringing in additional (non-CIC) partners. The University of California ("UC"), in particular, was heavily pursued:

We need to bring in the University of California because the CIC produces about 10% of the PhDs, and the University of California produces another 10% of the PhDs. If we've got 20% of PhD construction it will be very hard for the others not to join. Once the two biggest institutionalized players are in, we'll get there.

4.3.1. Appealing to New Partners

Gaining UC's commitment proved to be a significant challenge. UC and the California Digital Library ("CDL") were global leaders in large-scale digitization. They had a history of working with the Internet Archive and Open Content Alliance and partnering with members of industry including Microsoft, Yahoo!, the Sloan Foundation and others prior to joining Google's MDP. As a senior CDL administrator described, UC saw itself as "the intersection of the Venn diagram of digitization:"

We had a great sense of the big picture, of what people were working on, how far they were, what kind of challenges they had, how they were thinking about access and preservation. We really were in the center of the communication and social side of digitization efforts.

UC's institutional identity and self-positioning had a number of implications (positive and negative) with respect to the UMs initiative. Weighing in UM's favor, the CDL had experienced frustration over the lack of organizational infrastructure on some of its prior collaborations: "We were accomplishing digitization but we were not accomplishing the infrastructural aspects the libraries needed." UM and IU had a proven track record of successfully implementing collaborative and innovative projects. In addition, the CDL was concerned about a misalignment of values between the library community and private firms like Google, Microsoft, Yahoo! and others. A CDL representative noted:

The academy traditionally tries to solve problems like each one of us are an island but the digital favors scale. Either we figure out how to create scale ourselves in ways that we can steer in our interest, and take some advantage of the economics of it, or others will create scale and they will manage it in ways that are not in our interest.

The CDL saw value in building a DI by, for and of research libraries. The SDR would preserve, organize, and manage the data in a way that was consistent with library values and practices.

There were also a number of factors that weighed against the SDR from the perspective of UC and CDL. UC had intended to develop its own DI and progress was well underway when it was asked to consider abandoning it in favor of UM's which they viewed, unimpressed, as a regional CIC project. In addition, the UC system is particularly large and particularly bureaucratic. Reaching consensus among the twelve UC libraries, and between the libraries and Office of the President (UCOP), is perhaps even more daunting than reaching consensus among the CIC. When the CDL eventually approached UC's governing board advocating for UC to join the initiative, UC took no action. In the view key administrators at UM, the situation was getting dire:

It was easy to get the CDL people to join because this was right up their bailiwick. But it was clear to me from the start that this wasn't going to go anywhere unless we got Berkeley and UCLA on board. They are by far the biggest pieces of the UC system in terms of campuses and they have stopped things repeatedly in the past. If Berkeley and UCLA gang up they are essentially invincible. So, we didn't necessarily need them to say, 'We're in. We love it.' But we at least had to get them to say, 'We won't fight it.' That took about a year.

UC ultimately did decide to join and its rationale was twofold. One justification was economic — it was far more cost effective to share a single DI than create and support its own. A second justification dealt with salience and control. As UC sat on UM's invitation word began to trickle out that something big (HT) was about to be announced. If UC wanted the privileges afforded to founding members, i.e.

organizational power to shape HT, it had to join now. Again, UM essentially forced action with an implied ultimatum. As described by a senior administrator:

We are moving forward with or without you. If you join us now, we'll give you a seat at the table, but if you wait, you won't get that level of status within the organization.

As it did with the CIC, the bold exercise of power paid off in terms of positioning to DI to be a viable and robust offering for the research library community. UC joined and UM was able to push the SDR through a stagnating decision-making process. The shorter-term instrumental and partnership problems had been solved. Now its co-creators looked toward the long now of HT.

For HathiTrust to succeed over the longer term, its progenitors recognized that UM could not operate the repository as a dictatorship but must cede control over to the collective. As one of my participants described:

The library community is very catty. Because they've been deprived of power for so long they engage in horizontal violence at the local level. So, the number one complaint would be that Michigan is doing this thing that really benefits us so that they can control us. This was going to be a huge issue. And so we had to give HathiTrust over to the members of the community, so that they could settle upon what HathiTrust might become. We couldn't say 'This is the direction it's going to go' because, even if we were right, it would be prima facie evidence that we were drunk with power, and mad, and taking them where they didn't want to go. We had a vision, which was that we really needed to back-up our digital scans, but the rest had to be settled by the library community.

When HT was formally introduced to the public in the fall of 2008, it was announced as a SDR jointly founded by the 12-university consortium known as the CIC and the 11 libraries of the University of California system. There was no specific mention of UM or IU beyond the fact that they were members of the CIC. UM's institutional fingerprints were already fading from the HT creation story, enabling new meanings to emerge out of the new collective. A UM librarian observed:

When you're at Michigan, you see what's going on here. It wasn't until I was at a CIC meeting and saw people with HathiTrust stickers on their computers and heard them referring to HathiTrust as 'We' rather than as 'Michigan' that I realized there was already this broad sense of collective action being expressed around HathiTrust. It was really an amazing thing to see.

6. Conclusion

HT's dark history reveals the multiple, entangled, dynamic and sometimes unexpected ways that values, power, and politics shape and are shaped by the development of large-scale, collaborative digital infrastructures. Although much of its early history was founded upon unilateral and bilateral power plays that stood in stark opposition to traditional library values, practices, and governance structures, these moves were critical to HT's creation and fed into its success nearly a decade post-launch. This history also has implications for libraries and digital infrastructure more generally. HT's emergence and evolution privileges institutional partners that possess certain traits (i.e. large federated research library with significant English-language print collection) and have access to certain resources (i.e. funding and high quality, reliable broadband Internet) while effectively excluding other kinds of participants (i.e. individuals, smaller municipal libraries, private firms), and other forms of participation (i.e. membership without contribution and contribution without membership).

This descriptive account and analysis of a particular DI reveals some of the ways in which existing relationships and alliances, shared values and practices, organizational sensemaking and institutional structures may become inextricably bound up and entangled in DI development. The MDP enabled but did not lead to the HT we see today in a linear or deterministic sense. Instead, HT *became* over time, taking form through a process of incremental steps, unanticipated challenges and responses to changing technical, social, and institutional conditions. Values, power, and politics were implicated across a sensemaking and decision-making spectrum that ranged from overt ultimatum to strategic nudge to passive self-exile. Using descriptive retrospective accounts generated from participant interviews triangulated against a rich textual record this research contributes to a more complete picture of the ways values, social relationships, organizational strategy, and institutional politics influence digital infrastructure development. In particular, it demonstrates the value in foregrounding and emphasizing some of the hidden, subtle and more nuanced ways that values, power, and politics influence digital infrastructure development.

7. References

[1] Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological*

systems: New directions in the sociology and history of technology. MIT press.

[2] Bowker, G. C., Edwards, P. N., Jackson, S. J., & Knobel, C. P. (2010). 1.1 The Long Now of Cyberinfrastructure. *World Wide Research: Reshaping the Sciences and Humanities*, 40.

[3] Conway, P. (2013). Preserving imperfection: Assessing the incidence of digital imaging error in HathiTrust. *Preservation, Digital Technology & Culture*, *42*(1), 17-30.

[4] Conway, P. (2015). Digital transformations and the archival nature of surrogates. *Archival Science*, *15*(1), 51-69.

[5] Duguid, P. (2007). Inheritance and loss? A brief survey of Google Books. *First Monday*, *12*(8).

[6] Edwards, P. et al (2013). Knowledge infrastructures: Intellectual frameworks and research challenges.

[7] Grimmelmann, J. (2008). Google Dilemma, The. *NYL Sch. L. Rev.*, *53*, 939.

[8] Grimmelmann, J. (2011). The Elephantine Google Books Settlement. *Journal of the Copyright Society of the USA*, *58*, 497.

[9] HathiTrust Press Release, October 13, 2008, "Launch of HathiTrust – October 13, 2008."

[10] Hoffmann, A. L. (2014). Google Books as Infrastructure of In/justice: Towards a Sociotechnical Account of Rawlsian Justice, Information, and Technology.

[11] Hoffmann, A. L. (2016). Google Books, Libraries, and Self-Respect: Information Justice beyond Distributions. *The Library*, *86*(1).

[12] Hoffmann, A. L., & Bloom, R. (2016). Digitizing Books, Obscuring Women's Work: Google Books, Librarians, and Ideologies of Access. *Ada: A Journal of Gender, New Media, and Technology*, (9).

[13] Hughes, T. P. (1993). *Networks of Power: Electrification in Western Society*, 1880-1930. JHU Press.

[14] Jones, E. (2011, February). Large-scale book digitization in historical context: outlines of a comparison. In *Proceedings of the 2011 iConference* (pp. 829-830). ACM.

[15] Jones, E. A. (2014). *Constructing the Universal Library* (Doctoral dissertation, University of Washington).

[16] Kling, R. (Ed.). (1996). *Computerization and controversy: value conflicts and social choices*. Morgan Kaufmann.

[17] Leetaru, K. (2008). Mass book digitization: The deeper story of Google Books and the Open Content Alliance. *First Monday*, *13*(10).

[18] Ribes, D., & Finholt, T. A. (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, *10*(5), 375.

[19] Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information systems research*, *7*(1), 111-134.

[20] Vaidhyanathan, S. (2006). Googlization of Everything and the Future of Copyright, The. *UC DAViS l. reV.*, *40*, 1207.

[21] Vaidhyanathan, S. (2012). *The Googlization of everything:(and why we should worry)*. Univ of California Press.

[22] Waller, V. (2009). The relationship between public libraries and Google: Too much information. *First Monday*, *14*(9).

[23] Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Sage.

[24] Weick, K. E. (2012). *Making sense of the organization, Volume 2: The impermanent organization* (Vol. 2). John Wiley & Sons.